Monetizing Data Management

17 Case Studies Illustrating How Data Leveraging (Big and Small) Can Produce Quantifiable Results That Are of Keen Interest to C-Suite Occupants

by

Peter Aiken, PhD
Virginia Commonwealth University/Data Blueprint

with

Juanita Walton Billings, MS/IS
OfficeWorks

Table of Contents

1.1 LIST OF FIGURES

1.2   FOREWORD BY JOHN BOTTEGA

It seems that not too long ago, any discussion regarding data management would have invoked one of three reactions… (1) Eyes glazed over; (2) a faint recognition of the importance data management, but little support or investment; or (3) an assumption that data management is a tech-only challenge and should be passed over to technology. Today, things have changed. Businesses large and small, across all industries, are coming to grips with the need to recognize, prioritize and manage data, as one of the most fundamental business assets of their institutions.

What has changed? And what is yet to be understood?

We have seen a dramatic shift towards understanding data as a result of several key events. The financial crisis of 2008 was one such event and a turning point for data management discipline. Data was not the cause of the crisis, but not being able to aggregate, compare, and truly understand the data that was before us impacted our ability, or the ability of the key decision makers, to react and respond. Imagine the best alarm system in the world installed in your home or office, that when activated, shouted commands to you in a foreign language that you could not understand. All of the data provided by the alarm system would have been correct. The delivery of the data was on time and on spec. But the 'meaning' of the data was missing. And without meaning, without that understanding, your response to the data would have been flawed, resulting in potentially fatal outcomes. The financial crisis brought this reality home. Understanding our data was suddenly promoted from the back office, to the c-suite. Data practices were being discussed in industry, across government agencies, and built into regulations. Data, data standards, and data management could be heard in testimony to Congress! The knock-on effect of not-knowing what our data was telling us became real.

And while events such as this were taking place, data itself was growing. If not understanding our data was bad enough; this same data was multiplying at enormous rates. Systems capturing hundreds of thousands, millions, or billions of events or transactions, are flooding into business systems. Databases are growing, capacity is expanding, but is the know-how to 'understand' and make sense of all this data, keeping pace?

Fair to say, we now know that data is important. We now know that data is critical to our everyday activities. Understanding our data is critical to running our business processes more efficiently, to controlling our costs, to creating better and more innovative products, and to servicing our customers better. But do we really know how to realize the true value of data? Do we really know the monetary impact of our data management decisions (or lack thereof)?

The bottom line is – data is the new currency. Yes, an expression that is being used quite a bit of late, but it is very relevant in discussing the importance of data and the methodologies by which we manage it.   And like any currency, how we manage it determines its true value. Like any currency, it can be managed wisely, or it can be managed foolishly. It can be put to good use, or it can be squandered away. The

question is – what factors determine the path that we take? How do we properly manage this asset and realize its full value and potential?

In Monetizing Data Management, Peter and Juanita explore the question of how to understand and place tangible value on data and data management. They explore this question through a series of examples and real-world use cases to exemplify how the true value of data can be realized. They show how bringing together business and technology, and applying a data-centric forensic approach can turn massive amounts of data into the tools needed to improve business processes, reduce costs, and better serve the customer.

Data monetization is not about turning data into money. Instead, it's about taking information and turning it into opportunity. It's about the need to understand the real meaning of data in order to extract value from it. And it's about achieving this objective through a partnership with business and technology. Data management is not a discipline reserved for technology alone. Data management requires business understanding and an understanding of the subject matter at hand, coupled with innovative and supportive technology, to fully realize the true value of data – to inform, expose, and direct business activity to better serve its intended use. In Monetizing Data Management, the authors address these objectives and demonstrate how true value can be realized from our data through improved data centric approaches. Enjoy the read but make use of the examples …

John Bottega, noted industry and data management expert – September 2013

1.3 EXECUTIVE SUMMARY (AND BACK COVER MATERIAL)

We seriously doubt that anyone would have purchased this book if it were titled: *Leveraging Data Assets to Reduce Organizational Risk,* but that is really what this book is about. More practically, the real title of this short book should be: *Monetizing (and otherwise elevating the importance of) Data Management So It Will Get Some Much Needed Attention in the C-Suite.*

That title is too long because it violates all search engine optimization rules. So, knowing that most C-level conversations eventually turn to money, we opted for the shorter title. However, knowing that some organizations have other top-line motivations, we also included non-monetary examples where data management has saved lives and legal bills – both contributing to the concept of organizational risk.

The concept of data as an asset that should be used to support organizational strategy is not sufficiently well known to those whose job it has become to manage data and organizational assets in general. Supported by our approach to monetizing and otherwise elevating the importance of data management practices, you will recognize the importance of improving the data management within your enterprise and the value of leveraging well-managed data. Not only will you more fully understand how data can be used in innovative ways to achieve strategic objectives, you will also no doubt identify opportunities to eliminate wasted resources and build in productivity advantages.

This book defines the five interrelated best practices that comprise organizational data management, highlighting its importance in today's organizational environment and explaining why it is likely being accomplished less than optimally in your organization. Further, it describes a value-added approach to managing enterprise resources and assets—a data-centric method that provides opportunities for organizational improvements and inherent pay-offs.

The 17 specific cases (of varying lengths) we relate will help you to identify opportunities to introduce data management into the strategic conversations that occur in the C-suite. You will gain a new perspective regarding the stewardship of your data assets and insulate your operations from the chaos, losses and risks that result from traditional approaches to technological projects. And you will learn how to protect yourself from legal challenges resulting from outsourced information technology projects gone badly due to incorrect project sequencing and focus. With the emerging acceptance and adoption of revised performance standards, your organization will be better prepared to face the coming *big data deluge!*

1.4   INTRODUCTION

This little book emerged from a series of talks given around the world by the authors faced with the task of helping our data management community better articulate the importance of what we do for organizations. After due consideration and much consultation with our peers, the answer became obvious. Until the data management community can meaningfully communicate in monetary or other terms equally important to the C-suite, we will continue to struggle to articulate the value of its role within organizational evolution.

Today's business executives are smart, talented and experienced experts who have done incredible things with their organizations. Unfortunately, they have been taught by our educational system that data management is a technical skill that belongs in the bowels of information technology organizations. This results in executive decision-makers being far removed from day-to-day information, and especially data, operations and, for the most part, being insufficiently data knowledgeable to make good data decisions. Consequently, too many decisions about data have been poor.

We organized the book into four parts.

The first Part gives a somewhat unique perspective to the practice of leveraging data. We describe the motivations and delineate the specific challenges preventing most organizations from making substantial progress in this area.

Part 2 presents 11 cases where leveraging data has produced positive financial results that can be presented in language of immediate interest to C-level executives. To the degree possible, we have quantified the effect that data management has had in terms that will be meaningful to them also.

Part 3 describes five instances taken from the authors' experiences with various governmental defense departments. The lessons in this section however can be equally applied to many non-profit and non-defense governmental organizations.

Part 4 speaks specifically to the interaction of data management practices, in terms of both information technology projects and legal responsibilities. Reading it can help your organization avoid a number of perils, stay out of court and better vet contractors, experts and other helpers who play a role in organization information technology development.

This book is, we hope, the answer to a question posed by many following our lectures: "Can you get this information into a form where I can hand it to my boss and ask that it be read on the next plane ride?" So, please hand this to your boss, and let us know whether or not it works.

2 Part 1 – Data Management as a Prerequisite to Data Leveraging

Part 1 frames the context for data management by defining the five interrelated practices that comprise organizational data management, highlighting its importance in today's organizational environment, explaining why it is likely being accomplished poorly in your organization today and describing a better approach—a data-centric approach to technology initiatives that highlights organizational pay-offs. Part 1 concludes by introducing our approach to monetizing these practices.

2.1 WHAT IS DATA? AN OBJECTIVE DEFINITION

Data represents an organization's sole non-depletable, non-degrading, durable, strategic asset. You can't use it up. If properly maintained, it cannot degrade over time or from use. It is, by accounting definitions, durable–persisting beyond the one-year yardstick applied by standard accounting practices. Data's value increases as it evolves along the value chain. From business predications, it melds into transactions and, ultimately, returns full cycle–becoming the basis for future predications of organizational strategy. So far, we have failed to acknowledge data's primary potential value: fit-for-use factual information that describes an organization's operational environment and improves decision-making. When combined, these attributes render data assets unique within the organizational repertoire. Data assets demand to be managed as professionally and aggressively as any other company asset. Objective measurements show that few organizations achieve data management success, resulting in an inability (or at least great difficulty) to exploit a strategic data advantage. In the face of ongoing data explosion, this leaves most organizations unprepared to properly leverage their data assets.

Figure 1 shows that formal definitions of an organzation's data and information are necessary but insufficient prerequistes to organizational use of information as a strategic asset. The figure is an extension of Appleton's original model [see (Appleton 1986)], specifying the relationships and dependencies between the concepts of INTELLIGENCE, INFORMATION and DATA. The elegance of his original model derives from its pictorial explanation of a hierarchical dependency of an organization's strategic use of information on the information and data layers below. Each layer represents a necessary but insufficient prerequisite to organizational quests to leverage data. From the bottom:
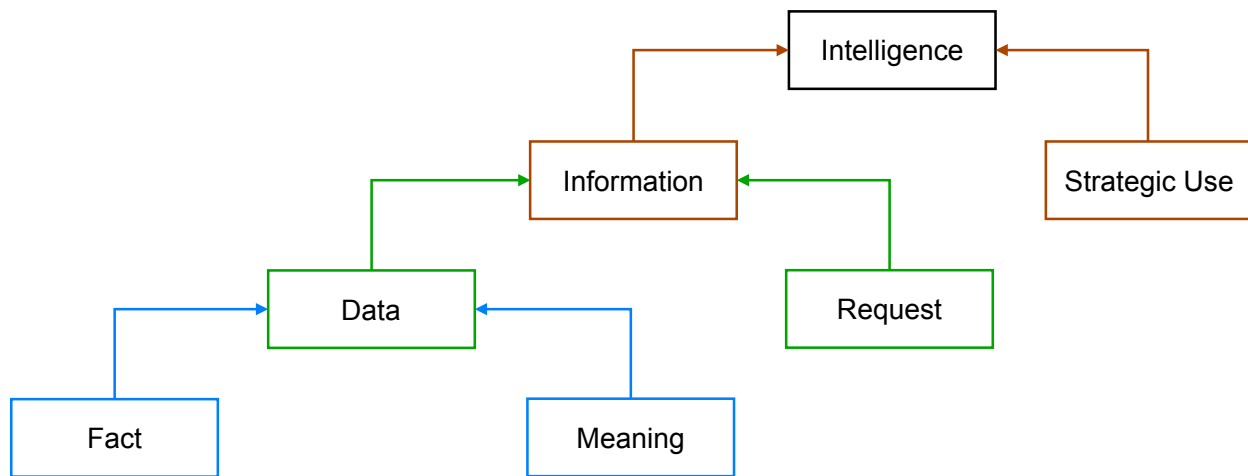
**Figure 1   A model showing how data and then information are necessary but insufficient
prerequisites to obtaining strategic information use**

1. DATA is a combination of a FACT explicitly paired with one or more MEANINGs.

    Each FACT combines with one or more MEANINGs.

    Each specific FACT and MEANING combination is referred to as a *datum*.

    Data is based on its discrete value and the definition of that value: fact and
    meaning. The formal arrangement of an organization's data assets is referred to
    as its *data architecture*. All organizations have data architectures; some
    organizational data architectures are better understood and therefore more
    useful than other organizations' data architectures.

2. INFORMATION, in turn, is based on the DATA layer and combines specific data
    with specific REQUESTs.

    An INFORMATION is one or more DATA that is returned in response to a
    specific REQUEST.

    Incorporation of the various requests is the only thing that differentiates an
    information architecture from a data architecture. Since the two are intrinsically
    interrelated, organizations must manage them as one asset—and refer to it as
    either its data architecture or its information architecture. Attempting to manage
    them separately causes far more harm than benefit. For this book, we will use
    the term *data architecture*. *You must select the more appropriate term for your
    organization and you must train everyone to use the selected term only—not
    both.* The confusion is not worth the cost.

3. *Information reuse* is enabled when one FACT is combined with more than one MEANING.

   Information reuse offers far more potential than software reuse – a concept that is largely seen as a failed endeavor.[1] This is attributed to currently unresolved issues with definitions, component classification and existing software development practices. NOTE: these same issues inhibit data reuse, but the challenges are orders of magnitude simpler. And data-focused methods are much more mature.

4. INTELLIGENCE[2] is derived from understanding both information and its associated frame of reference: its strategic use.

   INTELLIGENCE is INFORMATION associated with its STRATEGIC USEs.

   To make the most effective use of organizational data and information resources, they must be formally managed.

   DATA and INFORMATION must be formally arranged into an *architecture*.

   The organizational INTELLIGENCE layer must be developed using specification and common understanding of the subordinate data architecture.

As with any architectural construct, one cannot build a complex, lasting structure without a strong foundation. A multi-story house cannot be built upon a base of marshmallows. If the foundation is unable to support higher levels, the entire effort—and structure—will be crippled. The data and information layers are foundational building blocks.

Failure to understand these initial architectural concepts about the relationship between data, information and its strategic use are the primary reason so many organizational business intelligence/analytical initiatives fail or take longer, cost more, deliver less and present significantly increased organizational risk!

Without diving too deeply into the subject, it should now be obvious that creation of separate piles of business intelligence/analytical data is counter-productive in most instances – creating duplicate data and wasting effort by cleaning analytical data instead of reforming source data handling practices (to name just one common practice).

The rest of this part will address a number of topics somewhat familiar to data managers, but perhaps not as well known to organizational business executives— and especially your C-level executives. We will discuss:

---

[1] Google the phrase "how much software is really reused" and you will find hundreds of reports indicating that the cost of reusing software is now widely perceived as far more expense than recreating it or purchasing new software packages.

[2] The terms *wisdom* and *knowledge* are often used synonymously.

- the engineering and architectural concepts behind data leveraging;
- a formal definition of data management;
- the importance of data management to organizational success;
- the benefits of value-added data management versus traditional data management;
- the need to re-examine information technology support of the business from a data-centric perspective;
- our approach to monetizing data management;
- detailed examinations of five data management practice areas; and
- the critical need to improve organizational data management maturity.

2.2 DATA LEVERAGING REQUIRES BOTH ARCHITECTURAL AND ENGINEERING DISCIPLINES

Most refer to *management* as being a goal.[3] Management, however, is a necessary (but insufficient) prerequisite to successful data <u>leveraging</u>; leveraging is the higher-order objective. Advantageous data management is a prerequisite to achieving organizational data leverage, that is, strategic use of information. Exploiting a data advantage is what brings recognized value to data assets. Data assets must be perceived holistically and, more importantly, independent of technology—in the same manner in which a chief financial officer understands the available range of financial assets and instruments. The concept, *in support of organizational strategy*, is foreign to those whose job it has become *to manage data*.

Good data management practices must precede effective, innovative organizational data use and, specifically, information technology development projects. Failure to focus on this foundation ensures that value emerging at the intelligence level is diminished as a result of taking longer, costing more, delivering less reliable fact and presenting greater risk to the organization. Poor data architectural foundations represent a particular threat to the allure of big data leveraging techniques.

The process of building your data architecture must be managed inductively—based on factual information—derived from accurate and timely data that already exists. Such information can be effectively and efficiently extracted using reverse-engineering techniques that have been refined over decades [see (Aiken, Muntz *et al.* 1994)]. When combined with some exciting automated reverse-engineering technologies,[4] organizations can rapidly reclaim mastery of their data assets.

Similarly, without knowledge of the engineering disciplines, it is impossible to effectively leverage data within the organization and among partners. Leverage is achieved using data-centric technologies, processes and human skill sets. Leverage is increased as <u>r</u>edundant, <u>o</u>bsolete or <u>t</u>rivial (ROT) data is eliminated from organizational data (Figure 2). Perceiving and treating data as an asset simultaneously (a) lowers information technology costs and (b) increases knowledge worker productivity.

---

[3] It is an important goal, consider the question: "*How can you secure it if you can't manage it?*" and associated implications for the information security industry.

[4] See for example: http://globalids.com
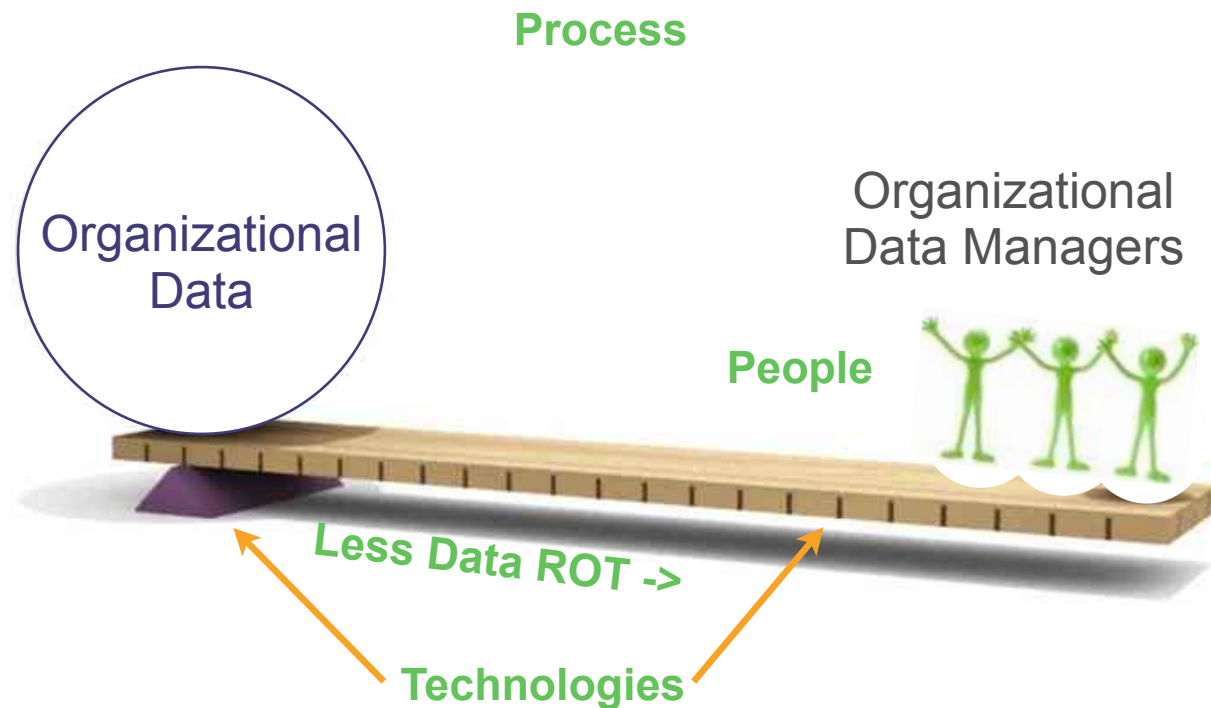
Figure 2   Leverage as an engineering concept – levering effectiveness is increased as data ROT is reduced

2.3   WHAT IS DATA MANAGEMENT?

Data management is generally a poorly understood concept from all perspectives. It is not taught as part of the vast majority of information technology programs, and therefore most information technology professionals and organizational leaders are *not data knowledgeable*. Since one cannot know what one has not learned, it is unfair to blame knowledge workers or executives for not knowing that data management is crucial to successful implementation of information strategies. Consequently, this condition represents a key constraint to comprehensive organizational strategy implementation. However, any holistic examination of the information technology field will reveal that it is largely about *technology*[5]—not about information. So, we can begin by stating that data management is largely about *putting the "I" back into IT*.

A good place to begin developing an understanding of data management is with the *Data Management Body of Knowledge*, often referred to in the industry as the *DM BoK*. Published by DAMA International[6] in 2009, it is the *de facto* guide to understanding data management – quoting:

> Data management is the business function (sic) of planning for, controlling and delivering data and information assets. This function (sic) includes the:

---

[5] Comprising 90% of scientific research and writing.

[6] Data Administration Management Association (dama.org) – DM BoK is available at amazon.com – keywords "DAMA DMBOK."

- disciplines of development, execution and supervision of
- plans, policies, programs, projects, processes, practices and procedures that
- control, protect, deliver and enhance the
- value of data and information assets.

Data management is known by many other terms.

- information management (IM)
- enterprise information management (ELM)
- enterprise data management (EDM)
- data resource management (DRM)
- information resource management (IRM)
- information asset management (LAM)

All these phrases are generally synonymous, but this writing will consistently refer to *data management*. Often the word *enterprise* is included in the phrase to emphasize the enterprise-wide focus of data management efforts, i.e., enterprise information management or e*nterprise data management*. Enterprise-wide data management is a recommended best practice. However, data management may also be performed effectively in a local context without an enterprise-wide scope or mandate, although with less business benefit.

Data management encompasses tasks commonly referred to as *database administration with database design, implementation and production support*, as well as *data administration*. The term *data administration* was once a popular way to vaguely refer to all the processes of data management—except database administration. However, as the data management process matures, its specific tasks are better understood. The data management process is important to enterprises of all sizes and purposes.

The scope of the data management process and the scale of its implementation vary widely with the size, means and experience of an organization. The nature of the data management process remains the same across organizations, even though implementation details differ widely. (DAMA-International 2009)

The *DM BoK* uses a wheel (Figure 3) to describe 10 data management processes. While this representation does not show that some tasks are optional and others are dependent upon each other, it does serve as a useful starting point for an ultimate understanding that (a) these 10 processes are not taught in the vast majority of educational contexts and, thus, (b) require that knowledge workers seek out alternate sources to learn about them. DAMA International is the data management professional organization dedicated to improving the skills—and utilization—of such professionals.

We can safely state that data management is concerned with the development and execution of architectures, policies, practices and procedures in order to manage the information lifecycle needs of an enterprise in an effective manner. Data lifecycle management is a policy-based approach to managing the flow of an information system's data throughout its lifecycle—from creation and initial storage to that time when it becomes obsolete and is deleted. Effective data management involves well-thought-out procedures and adherence to best practices.[7] Further, it is the

---

[7] Tech Target at http://searchdatamanagement.techtarget.com/definition/data-management

administrative process by which required data is acquired, validated, stored, protected and processed and by which its accessibility, reliability and timeliness is ensured to satisfy the needs of the data user.[8]
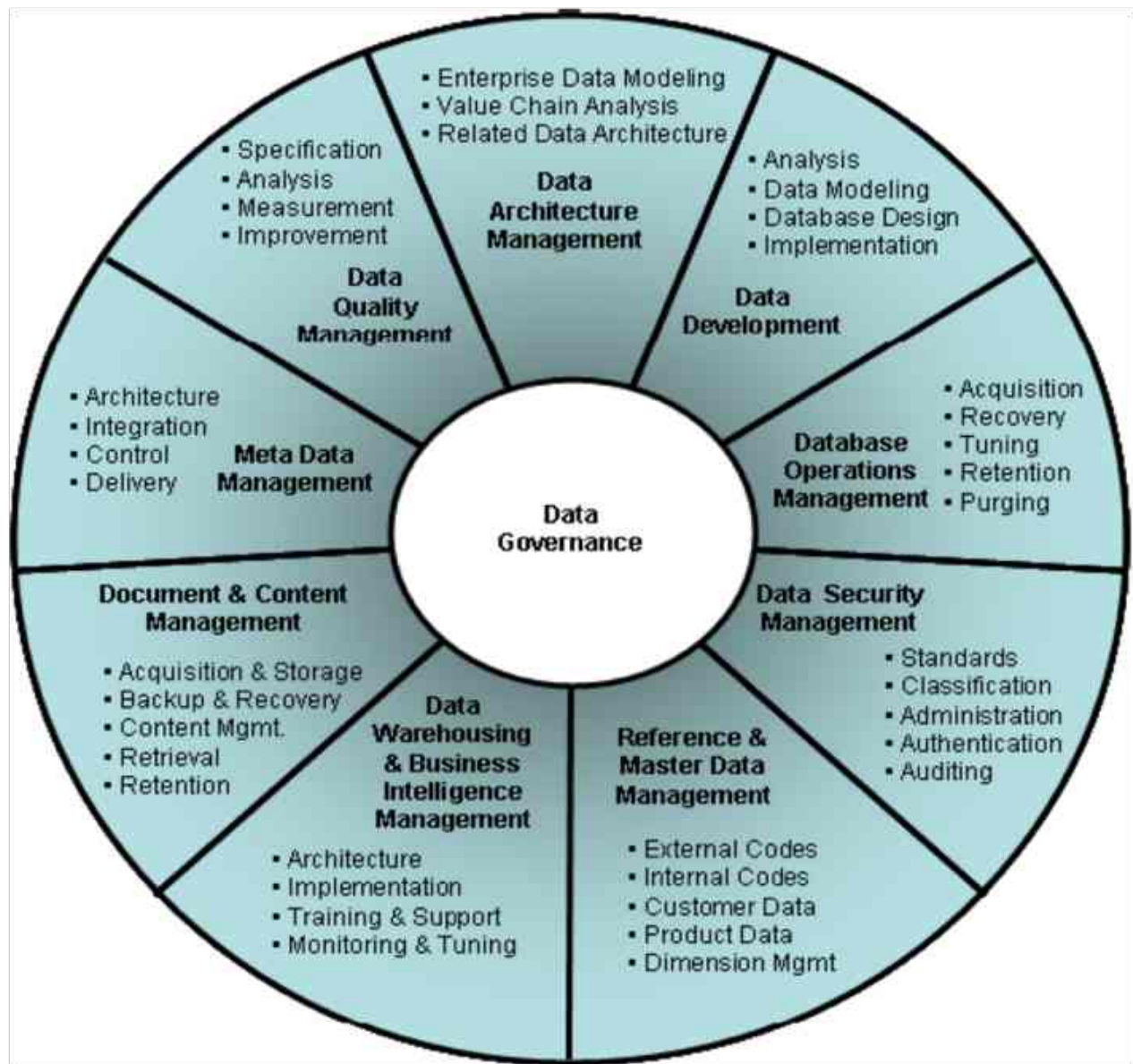
Figure 3   DAMA DM BoK Functions

2.4   WHY IS DATA MANAGEMENT IMPORTANT?

Data management improves organizational effectiveness in two primary ways.

2.4.1   Too much data leads directly to wasted productivity

The primary justification for data management is the fact that eighty percent (80%) of organizational data is redundant, obsolete or trivial (ROT). By taking excessive time

---

[8] Business Dictionary at http://www.businessdictionary.com/definition/data-management.html

to locate data needed to answer specific questions and by negatively impacting information reliability (fitness for use) as a result of poor data quality and calcification resulting from slow response to environmental changes, data ROT impedes and diminishes organizational performance.

2.4.2   Underutilized data leads directly to poorly leveraged organizational resources

Corporate resources consist of the means to satisfy customer demand for acceptable products and services in such a manner as to generate revenues and profits to sustain the business of the enterprise. More simply stated, corporate resources consist of the following high-level elements, all of which are dependent upon supply and managed by effectiveness and cost.

- **manpower** – costs associated with labor resources and market share
- **money** – costs associated with management of financial resources
- **methods** – costs associated with operational processes and product delivery
- **machines** – costs associated with hardware, software applications and data to enhance production capability

Manpower (labor) is traditionally the largest cost factor associated with producing a product or service. Often, however, it's the machines that unnecessarily consume a lion's share of the budget. And although many organizations have begun to recognize corporately owned data as an asset, several factors, both direct and indirect, can be blamed for the fact that it continues to be underutilized. More importantly, as a result of the situation, the corporate bottom line continues to feel an unacceptable negative impact. To help you understand, visualize an application of the organizational data management process.

Picture the process of planning a new restaurant. One alternative would be to go with the concept that presentation is everything. Each menu item would be served to patrons on a uniquely crafted plate. The apple pie plate might be shaped like an apple while the banana ice cream dish would be banana shaped. As a strategy, this might well reinforce the new restaurant's uniqueness to its customers. However, if the banana ice cream dish was broken during delivery, the cost of locating a new banana shaped dish might be high when compared to the alternative of serving everything using the standard plates. The cost of maintaining an inventory consisting of a large number of different serving plates might be high and would be factored into the planning, as would the costs of organizing the serving dish racks for easy access. Had the restaurant planners initially decided that (a) all plates will be the exact same size and all bowls will be the exact same size and (b) that all dinnerware will be the exact same color, these additional costs would have been avoided.

Unfortunately, organizations more often fail to include the cost of maintaining multiple, duplicative data items as part of their strategic calculations. When considering that our research shows that the average organization maintains

customer information in 17 source systems,[9] it becomes obvious that both IT and business management fail to give this concept (data planning) the same level of attention that is does other strategic variables. Our research indicates that this can cost organizations up to 40% of total IT costs.

2.4.3   An incorrect educational focus

We have been teaching smart people the wrong things for decades now. This strong statement means that the last three generations of STEM,[10] business and information technology leaders have been taught the wrong stuff about data and data management and, as a result, they are simply not qualified to make important decisions about organizational data assets and related matters.

Curricula encompassing computer science, information systems and computer engineering—and STEM in general—typically emphasize how to build new systems. As it turns out, this is a fundamental needs mismatch, especially when it is well understood that 80% of IT costs are spent rebuilding and evolving existing systems and only 20% of costs are spent building and acquiring new systems. The thought of putting fresh graduates on new projects makes this proposition more ridiculous. Clearly, given the poor IT implementation record, only the most experienced professionals should be allowed to participate in new systems development. Further, educators should be redirected to focus on the real challenge being faced by organizations—that being the evolution of existing data and systems architectures to better support organizational strategy.

Instead of approaching data management as a purely technical discipline, it is also critical to understand it has social (i.e., enterprise-wide) impacts as well. Therefore, it requires a blended social-technical approach. Unfortunately, an understanding of the necessary blended approach is sorely lacking on the part of those who teach it.

2.4.4   Lack of agreement over who is responsible for data assets

The information technology and business sectors do not agree regarding the various responsibilities related to data assets. The business sector assumes that data is being competently managed by the IT sector; whereas the IT sector is equally certain that the business sector is responsible. Further, disagreement exists within the IT sector, where it is commonly thought that data should fit into existing systems development lifecycle (SDLC) methods. Development lifecycles create something from nothing, whereas successful organizational data evolution must precede <u>any</u> software development lifecycle activity.

> *Data evolution is separate from, external to and must precede system development life cycle activities!*

---

[9]   The number 17 is a useful benchmark – if your organization maintains customer location in less than 17 disparate locations, you should report to the C-level that you are above average and ask them if they think that this is the correct number of places to maintain their organizational customer information!

[10]   STEM stands for science, technology, engineering and math.

Organizations achieve little success when attempting to <u>evolve</u> system with methods designed to <u>create</u> something from nothing. Each activity involves a fundamentally different focus.

If we were in charge of data management, our division of responsibility would be organized with all current data management functions (Figure 3) except for two, DATA DEVELOPMENT and DATABASE OPERATIONS MANAGEMENT, moving out of the IT sector (which manages *machines* and their attendant functions) into the business sector (which establishes *methods* and their attendant activities). Two business data management tasks, METADATA MANAGEMENT and DATA SECURITY MANAGEMENT, are explicitly shared with IT operational activities. Historically, these two activities report to IT. This realignment represents a needed, radical shift in thinking and allocation of resources. This argument is articulated more fully in *The Case for the Chief Data Officer* [see (Aiken and Gorman 2013)].[11]

2.5 VALUE-ADDED DATA MANAGEMENT IS DERIVED FROM DATA-CENTRIC DEVELOPMENT BEST PRACTICES

Using two figures developed initially by a colleague,[12] we differentiate between traditional *application-centric* and *data-centric* development strategies in the two figures below. The first (Figure 4) illustrates application-centric development—the strategy used by a vast majority of organizations.

Strategy

⇩

Goals/ Objectives

⇩

Systems/ Applications

⇩

Network/ Infrastructure

⇩

Data/ Information
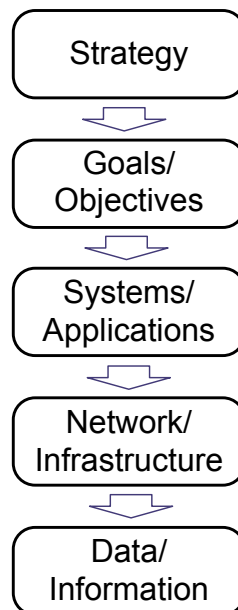
Figure 4   Application-centric development practices

---

[11] amazon.com keyword search "AIKEN CDO."

[12] We are indebted to Douglas Bagley (dougbagley73@gmail.com) for this specific conceptualization.

From the top:

- In support of an identified strategy (for example: *Make insanely great products!)*, the organization develops specific goals and objectives intended to achieve desired outcomes (for example: *Identify underserved market segments and develop products and services for them)*.

- Next, goals and objectives drive a configuration of specific systems and software applications (for example: *implement operating system "X" installed on server "Y" and purchase software package "Z")*.

- Implementation of those systems and applications leads to network and infrastructure requirements designed to support the configuration (for example: *implement software package "X" using middleware "Y")*.

- Data and information are typically only considered after other design specifications have been articulated (for example: *What data does software package "X" require?)*.

Unfortunately, challenges and drawbacks identified with this approach continue to permeate the combined IT/business environment. Processes and data are tightly coupled with applications, making it awkward and costly to maintain or change either the software or the data. Very little data reuse is available with an application-specific focus, and data integrity is at risk. Instead of accommodating organizational information requirements, data requirements are, by necessity, constrained by application requirements.

Figure 5   Data-centric development practices differ fundamentally from application-centric development practices

Figure 5, however, illustrates the fundamental shift in thinking required to, first, understand and, then, adopt data-centric development practices. Supporting the

notion that business needs (operational processes) should drive system design, the first two steps are common to both figures: determining organizational strategy and specifying specific goals and objectives intended to achieve the strategy. Next, ho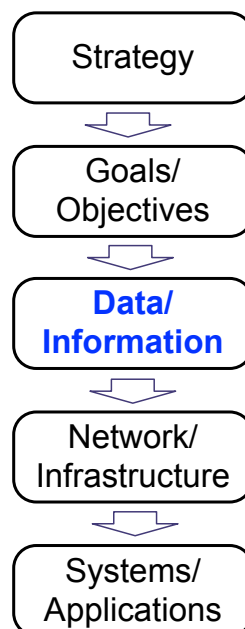wever, a description of information required to achieve measurable goals and objectives is derived from a data architecture that supports organization-wide data usage (for example: *Identify the characteristics, access channels, purchasing power for segments of the product buying public in excess of 100 million potential customers)*.

The identified data requirements should be the only variable when determining infrastructure, systems and software package requirements; all other network, software and/or application requirements, then, would be subordinated with delivery platform selection occurring last. In this manner systems and software can be specified and delivered that create the smallest possible footprint while focusing on previously articulated business goals expressed in terms of an organizational data model.

A recent implementation in which we participated involved migration of large transportation organization from its mainframe because of high costs and structural inflexibility. The organizational decision to adopt our data-centric approach led to an insight that all transactional processing could be done via mobile applications. In turn, this simplified the networking layer to a relatively simple HTTPS implementation. All that was left was a hugely simplified re-skilling of the existing, very competent IT applications group for secure, mobile platform delivery via tablets and the costs of a multi-million dollar implementation was reduced by an order of magnitude!

2.6 DATA-CENTRIC APPROACH LEADS DIRECTLY TO ORGANIZATIONAL PRODUCTIVITY ADVANTAGES

When data assets are developed from an organization-wide perspective, systems support organizational data needs complement organizational process flows. Data reuse is maximized. *Situation brittleness*—the ease of breaking a hard-coded condition—can be eliminated as a result of separating processes and the data via development of an information architecture. The ability to share and maintain data*, particularly in cases where data is shared across functional areas*, is enhanced. A secondary benefit of shared data is enhancement of data integrity—the accuracy, completeness, timeliness and appropriateness of data. But, because a data-centric approach represents such a fundamental change in organizational thinking, it is generally not on anyone's list of things to monitor. Again, *they do not know what they do not know!*

2.7 DATA MANAGEMENT CONSISTS OF FIVE INTEGRATED PRACTICE AREAS
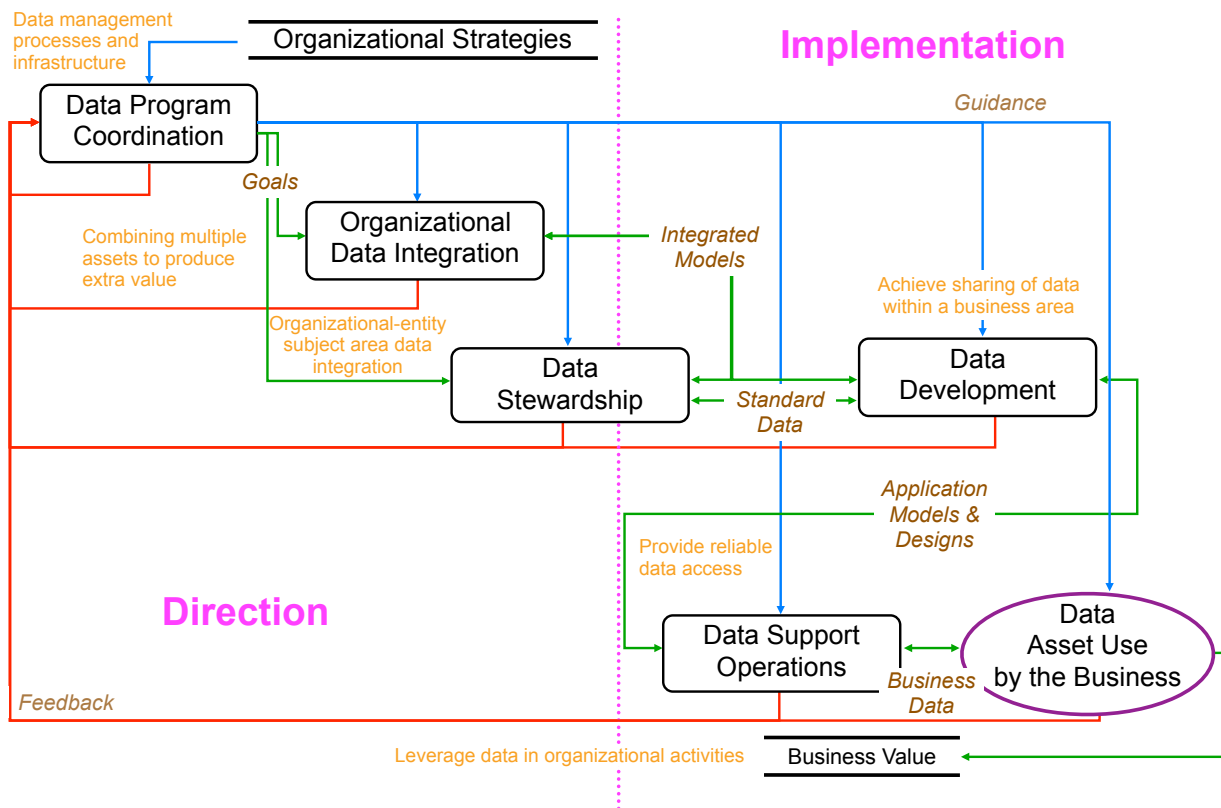
Figure 6   Five integrated data management practices

One of the most fundamental missing pieces of knowledge is the actual organization of data management practices. Data management consists of five interrelated specific practice areas (listed and discussed below). Figure 6 shows these as five rounded rectangles and how each is related to the others.[13]

- **Data Program Coordination** – data management practiced as a coherent and coordinated set of activities aimed at maximizing organizational benefit, specifically, defining, coordinating, resourcing, implementing and monitoring organizational data program strategies, policies, plans etc. as a coherent set of activities.

- **Organizational Data Integration** – delivery of data in support of organizational objectives, specifically, identifying, modeling, coordinating, organizing, distributing and designing data shared across business areas or organizational boundaries.

- **Data Stewardship** – designation of specific individuals as caretakers for specific data, specifically, ensuring that specific individuals are assigned responsibility for the maintenance of specific data as organizational assets, and that those

---

[13] See for more detail.

individuals possess the requisite knowledge, skills and abilities to accomplish these goals in conjunction with other data stewards in the organization.

- **Data Development** – efficient delivery of data via appropriate channels, specifically, specifying and designing appropriately designed data assets that are engineered to be capable of supporting organizational needs.

- **Data Support Operations** – ensuring reliable access to data, specifically, initiation, operation, tuning, maintenance, backup/recovery, archiving and disposal of data assets in support of organizational activities.

2.8   IMPROVING ORGANIZATIONAL DATA MANAGEMENT MATURITY: A FRAMEWORK

To date, organizations have not proven particularly competent at implementing information technology solutions; minimally, one in three projects (33%) is challenged on price, schedule and ultimate functionality. With more than 25 years of experience in the data management practice, we have observed that the premier shortcoming of organizations conducting information technology assessments and integrations has been a lack of willingness to invest in data management. We have not investigated an IT project failure that didn't ultimately uncover poor data management practices as the root cause of the failure.

Each data management practice area can be objectively quantified using the familiar Carnegie-Mellon University Software Engineering Institute (CMU/SEI) Capability Maturity Model (CMM) scale where scores, increasing from 1 to 5, represent the stage or condition of an organization's data management practices (see Figure 7). The CMM levels are:

1. Initial: Our data management practices are ad hoc and dependent upon "heroes" and heroic efforts.

2. Repeatable: We have data management experience and have the ability to implement disciplined processes.

3. Documented: We have standardized data management practices so that all in the organization can perform it with uniform quality.

4. Managed: We manage our data management processes so that the whole organization can follow our standard data management guidance.

5. Optimizing: We have a process for improving our data management capabilities.

Confidential Draft Version: *Monetizing Data Management-17* Case Studies Illustrating How Data Leveraging (Big and Small) Can Produce Quantifiable Results That Are of Keen Interest to C-Suite Occupants

data blueprint

We have a process for **improving** our DM capabilities

We **manage** our DM processes so that the whole organization can follow our standard DM guidance

We have experience that we have **standardized** so that all in the organization can follow it

Optimizing (5)

Managed (4)

One concept for process improvement, others include:
- Norton Stage Theory
- TQM
- TQdM
- TDQM
- ISO 9000

and focus on understanding current processes and determining where to make improvements.

Defined (3)

Repeatable (2)

We have DM experience and have the ability to implement **disciplined** processes

Initial (1)

Our DM practices are **ad hoc** and dependent upon "heroes" and heroic efforts

Figure 7   CMM levels have directly influenced such popular improvement frameworks as TQM and ISO 9000

A full assessment includes an industry comparison. Figure 8 shows a comparison performed for one of our clients in the airline industry. Scores for the airline that commissioned the comparison are indicated by the top bars of each group. A "weakest link" scoring system ensures that, since an engineered system can only be as strong as its weakest component, the airline received a score of 1 (the score of the weakest link) based on this assessment.

Increasing CMM Levels →

Figure 8   DM Practices within and across industry comparison

This type of analysis also indicates that a specific industry, the airline industry (middle bars), was ahead of the general state of the practice (the lowest bars in each group) in some practice areas and behind it in other practice areas. This parti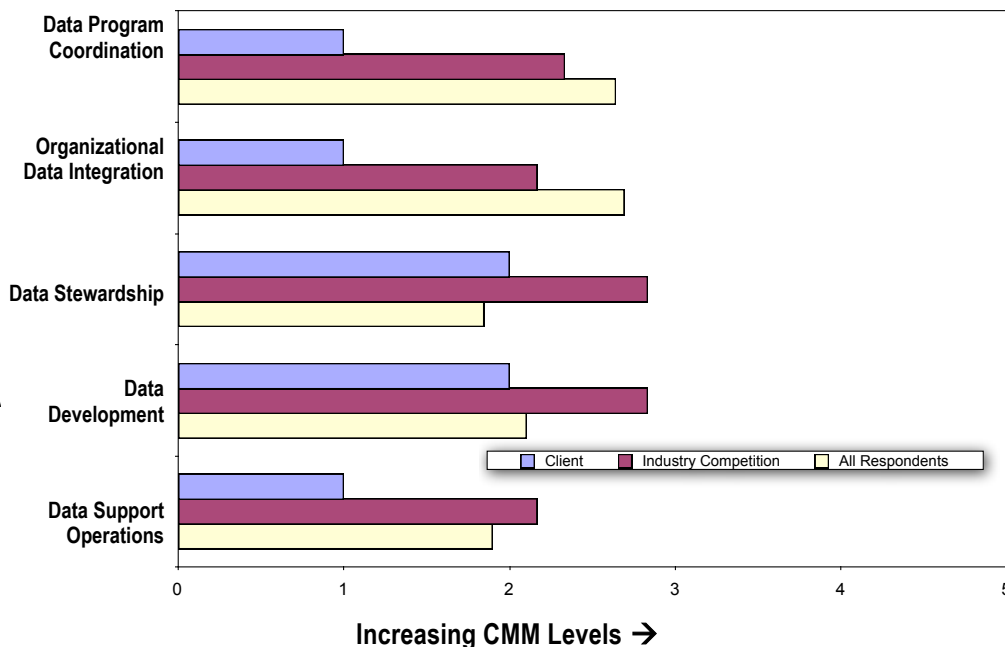cular airline's data management practices were noticeably less mature in the areas of *data program coordination*, *organizational data integration* and *data support*, especially when compared with other airlines that had been assessed.

These results are an extension of the original SEI CMM model. By increasing granularity of the model, organizations can interpret it as a roadmap indicating that, in this instance, the organization should take steps to enhance maturity of its data management practices by improving its data program coordination, organizational data integration, and data support operations—before attempting to improve the other two practice areas—as these investments will yield a greater ROI than the others.
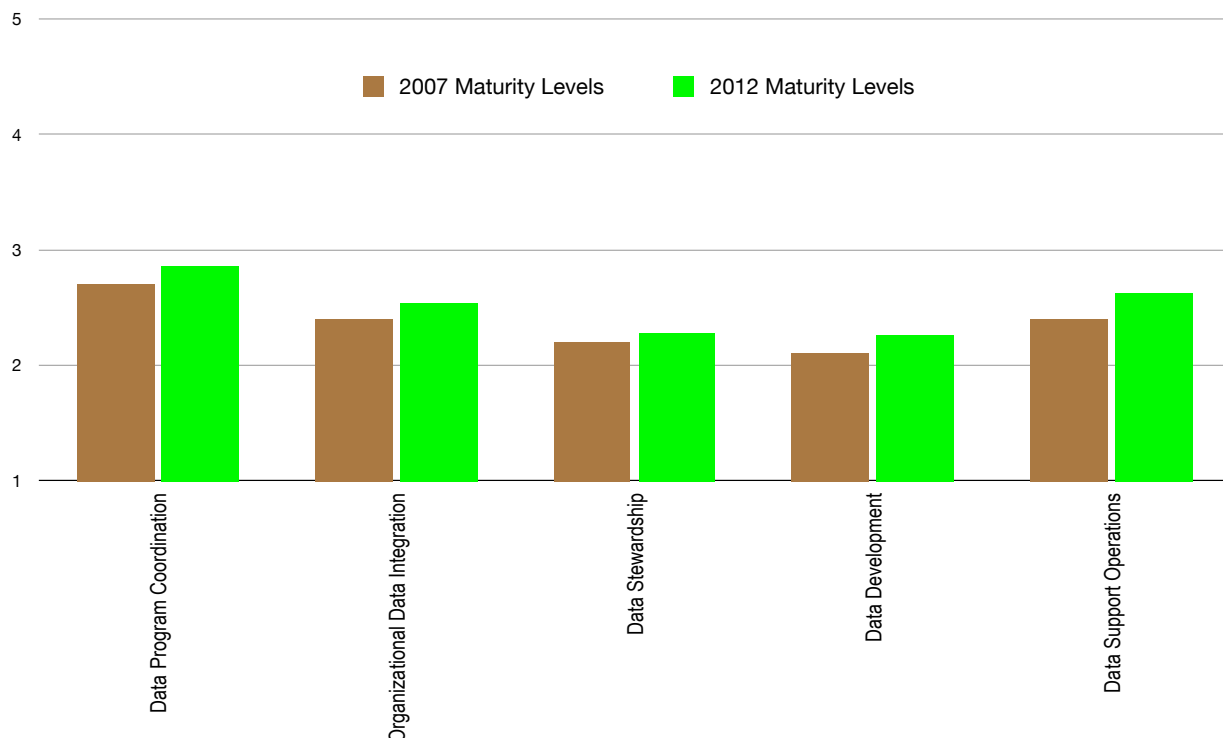
Figure 9   Comparison of industry wide data management maturity levels since dawning of the era of big data techniques (Aiken, Gillenson et al. 2011)

The objective scores (1-5) shown in Figure 9 can be used to benchmark and monitor industry-wide organizational data management efforts. We have reported on more than 500 organizational surveys and robust empirical measurements that detail the widespread state of data management and the numbers have not statistically varied in this era of big data. Research indicates that, on a scale of 1 to 5 (1 being poor, 5 being outstanding), most organizations are closer to one than to five. This indicates that the average organization *does not employ repeatable data management practice* (Aiken, Allen et al. 2007).

Confidential Draft Version: *Monetizing Data Management-17* Case Studies Illustrating How Data Leveraging (Big and Small) Can Produce Quantifiable Results That Are of Keen Interest to C-Suite Occupants

data blueprint

Consider the way children benefit from the retelling of stories as a pleasurable pastime, albeit one which builds memory structures. There is a reason that we crawl, walk, run—in that order. While we can't certify that organizations learn in the same manner as humans, there is a parallel. Until an organization gets good at data management, it cannot hope to achieve acceptable results from—or sufficient proficiency at—conducting advanced data practices such as predictive analytics or even basic data mining.

This is best understood by comparing data management practices to Maslow's hierarchy of needs. As shown in Figure 10, advanced data practices occupy the top of the hierarchy. We call them *silver bullets* because they are sold to corporate chief officers (referred to as *C-levels*) as solutions that will solve their problems. (As in, just purchase a solution in the green triangle of Figure 10 and your data problems will vanish.) Most organizations install these solutions with the same competence that they bring to other information technology solutions—that is to say, not much. So a poorly implemented system is then operated by an organization not equipped with sufficient knowledge, skills or abilities. Typically, these "solutions" become shelf ware or otherwise fail to deliver promised benefits.

Advanced
Data
Practices
• MDM
• Mining
• Big Data
• Analytics
• Warehousing
• SOA

Basic Data Management Practices
– Data Program Management
– Organizational Data Integration
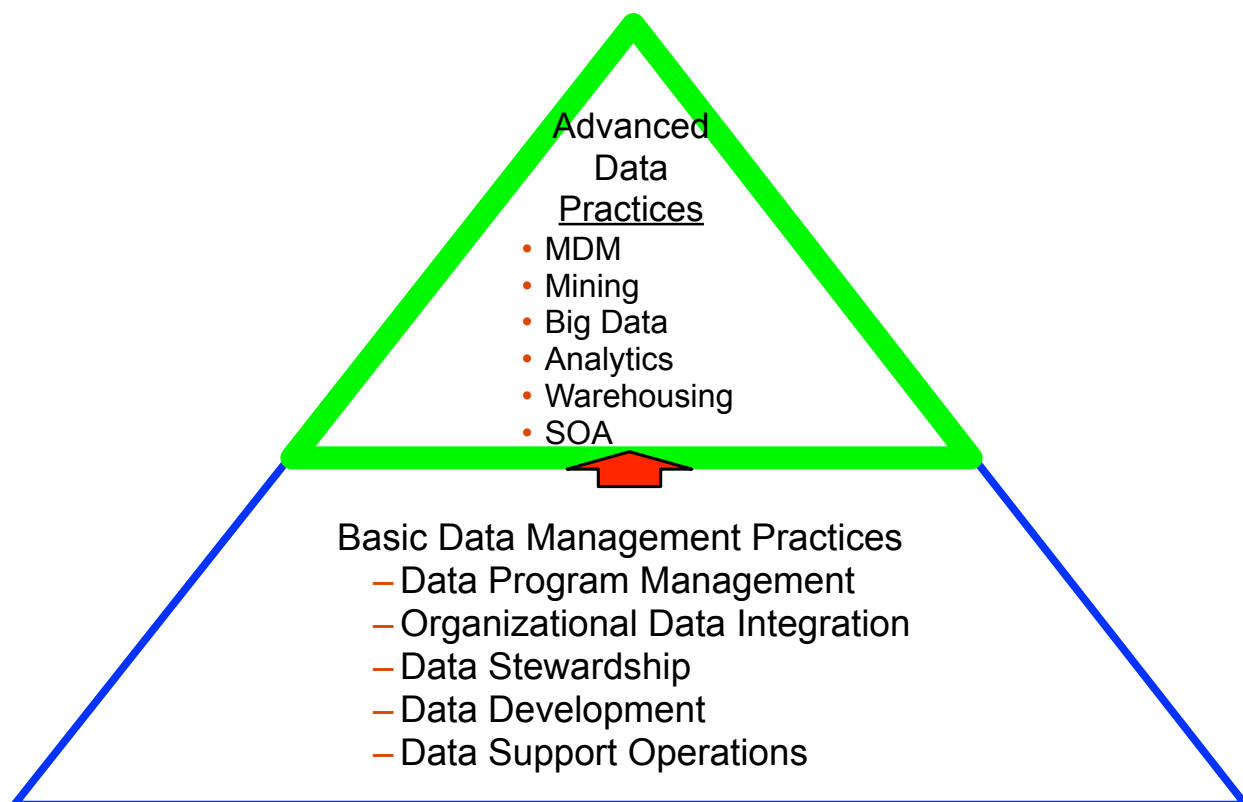– Data Stewardship
– Data Development
– Data Support Operations

Figure 10   DM Basics prevent organizations from spending too much, experiencing delays, falling short of objectives, and experiencing increased risk.

Most organizations fail at data endeavors because they attempt the really cool stuff with data—such as advanced analytics, business intelligence, data warehousing, master data management, big data (the list goes on and on)—before they gain proficiency in the five basic data management practice areas. If you are developing a data warehouse, or any other data-centric information technology system, and if

you have not gotten the basics down, it will cost more, take longer and the organization will incur greater risk. Bottom line: does it make sense that the *cost of correction should exceed the cost of the adjustment*?

2.9 DATA MANAGEMENT PAY-OFFS

Improving data management practices and the resulting capabilities adds value not only to the information produced, but also to overall organizational goals. Widespread benefits accrue to the entire organization by:

1. helping organizations prepare for future change by implementing a flexible and adaptable organizational data architecture;

2. focusing data assets to efficiently and effectively support organizational strategy;

3. increasing the percentage of time available to accomplish this by lowering the percentage of time required for maintenance activities;

4. reducing organizational data ROT;

5. permitting remaining data to receive more attention with respect to quality, security and reuse;

6. reducing the amount and complexity of the organizational code-base;

7. reducing the amount of time, effort and risk associated with information technology projects;

8. engineering flexibility and adaptability into data architectures instead of attempting to retrofit them after they are in production;

9. producing more, reusable data-focused work products;

10. permitting organizations, when faced with a choice between chaos versus understanding, to gravitate toward a cheaper, more understandable solution;

11. permitting organizations, when faced with a choice between complexity versus ease of implementation, to gravitate toward a cheaper, more understandable solution;

12. decreasing time spent understanding versus time spent considering the data-focused portions of organizational strategy; and

13. minimizing uncertain benefits versus engineerable benefits.

BOTTOM LINE: *Data management will, at a minimum, positively impact organizational resources* with regard to:

(a) solution development costs related to operations and maintenance time and efficacy,

(b) labor resources required for trouble-shooting and rework,

(c) improved data quality and resultant decision-making accuracy and speed, and

(d) decreased risk associated with untimely, misguided information.

2.10   MONETIZING DATA MANAGEMENT: AN INTRODUCTION

Unfortunately, sometimes, about the best that can be done is to estimate (and usually, under-estimate) the cost of individual participation in operational processes. That said, this nevertheless permits incorporation of a wider variety of cost modeling and remedial techniques within analysis efforts. Activity-based costing is just one of many measurement techniques available to evaluate data when communicating the leverage possible with advantageous data management. (For further information about this subject, we invite you to read *How to Measure Anything: Finding the Value of Intangibles in Business* by Douglas W. Hubbard, ISBN: 0470539399.) Personally, we were introduced to these techniques when working for and with the Defense Information Systems Agency's Center for Information Management in the early 1990s. Techniques presented in the book were used to approximate various costs and benefits associated with specific information technology investment decisions. More importantly, additional benefit accrued from the focused dialog addressing concrete concepts for increasing the ability to meet organizational needs.

To illustrate the absolute basics of activity-based costing alone will enhance understanding as to how to quantify business value. Let's look at the following four panels (Figure 11). The entertaining articulation below was created and shared but not credited by someone known to us.

Figure 11   Example of activity based costing (source unknown)

A better understanding of costs associated with the execution of specific tasks (processes), the interactions of the tasks (process design), the automation supporting the tasks (processing) and maintenance of operational models (process maps) will accrue from analysis. The same analysis can also give rise to improvements to each of the cost factors.

For instance, in Sheena's story the tasks are:

- driving distractedly,
- traveling through airport security, and
- pulling luggage from the overhead rack.

This tongue-in-cheek example illustrates an accounting of Sheena's time. It can be valued by superimposing an approximate salary; let's say $40 thousand per year. The impact of Sheena's tasks having been categorized as part of the critical path—and thus impacting others whose time can be totaled and valued as well—is illustrated. This permits broader claims such as "killing a whole person(-day)." According to Wikipedia, activity-based costing "... assigns more indirect costs (overhead) into direct costs compared to conventional costing models." So basically, by totaling the cost of a knowledge worker's time expended, we can approach some—but clearly not all—costs. This will permit statements such as "at least amount x is being spent by this organization for task y."

We have used this basic approach through for the past 20 years to calculate the minimal costs associated with poor data implementation and presented our calculations in the various cases presented in this book to illustrate to management the tangible costs associated with these decisions.

3   Part 2 – Bottom Line Pay-offs: Eleven Financial Cases

Part 2 describes eleven cases with specific monetary results illustrating how the practice of monetizing data management directly helped a number of organizations in tangible ways that were easy to relate to those in the C-suites. After all, these folks are smart and, when you bring them a plan with good return on investment and results, the next question is usually: where do we go next?

3.1   HOW MANY TIMES DO WE HAVE TO SPEND THAT MONEY? ($10 MILLION ANNUALLY)

Figure 12, a straightforward depiction of the pay grade, location and number of associates involved in time and leave tracking at a state agency, became locally famous for a couple of years. Shown on the Y-axis is the number of employees at a particular pay grade (on the Z axis) by location, in this case, district, (depicted on the X axis). Using the diagram, it is possible to illustrate the veracity of the statement that at least 300 individuals spent at least 15 minutes each week performing data management tasks associated with time and leave tracking for the agency.